View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Current Biology Vol 16 No 6
R186

Primer

Directed DNA deamination by AID/APOBEC3 in immunity

Donna A. MacDuff and Reuben S. Harris

Several hundred papers in the past seven years have established two new mechanisms of immunity, both of which are dependent upon proteins that deaminate cytosines to uracils within single-stranded DNA. Activation-induced deaminase (AID) deaminates C -> U to implement the diversification of vertebrate antibody genes and mammalian APOBEC3 proteins, such as human APOBEC3F and APOBEC3G, deaminate C -> U to trigger the destruction of a wide variety of retroelements including HIV. Here, we will discuss how the purposeful deamination of DNA-based cytosines underpins integral parts of both the adaptive and the innate immune responses.

Antibody diversity

The generation of a large and diverse antibody repertoire is fundamental to our ability to respond to and neutralize invading pathogens. Antibody diversity is achieved by three distinct events that irreversibly alter the immunoglobulin gene DNA specifically in the B-cell lineage. The first process (which also generates functional T-cell receptor genes) is V(D)J recombination. Functional antigen receptor genes are assembled in a site-specific recombination reaction catalyzed by the RAG1 and RAG2 enzymes, which juxtapose the germline-encoded variable (V), diversity (D) and joining (J) region DNA segments. Individuals that are unable to complete V(D)J recombination lack mature B and T lymphocytes, resulting in severe combined immunodeficiency.

B cells that have successfully completed V(D)J recombination migrate to lymphoid tissues where they encounter specific antigens. These cells express surface immunoglobulin of the μ isotype (IgM) and secrete soluble antibodies with relatively low affinities. Co-stimulation of B cells by antigen and a specific T cell triggers further rounds of diversification via somatic hypermutation (SHM) and class switch recombination (CSR). SHM introduces non-templated point mutations into the antibody gene variable region. Such mutations occur at an extremely high frequency and can alter the affinity of the encoded antibodies without changing their effector function. Iterative rounds of SHM coupled to antigen-dependent selection are responsible for antibody affinity maturation. CSR is a switch-region-specific recombination reaction that replaces the μ constant region with one of the downstream constant regions, γ, α or ε, thereby changing the antibody isotype to IgG, IgA or IgE, respectively (e.g. recombination of Sμ and Sγ in Figure 1). These alternative antibody isotypes mediate a range of activities, including antigen neutralization and marking, gut homeostasis and allergic responses.

AID

Activation-induced deaminase (AID) was cloned by Honjo and coworkers in 1999 as a factor that was specifically upregulated in B lymphocytes induced to undergo CSR. Its similarity to the *APOB* mRNA editing catalytic protein 1 (APOBEC1), which catalyzes specific mRNA C -> U deamination, suggested that AID might function in a similar manner. The interest in AID increased substantially when it was shown to be essential for CSR and, remarkably, also for SHM in mice and humans. Moreover, a seemingly unrelated mechanism of antibody gene diversification that occurs in birds and some other vertebrates, immunoglobulin gene conversion (IGC), was also shown to require AID. These data indicated that the yet-to-be-defined mechanism would be as conserved though the entire vertebrate lineage as the antibody response itself.

The surprising requirement for AID in all three antibody gene diversification processes, coupled with the fact that SHM is biased toward C/G base pairs in a variety of systems, helped lead Neuberger and colleagues to propose a unified model for antibody gene diversification, suggesting that all of these apparently distinct events are triggered by AID through the direct deamination of cytosines in DNA (Figure 1). In support of such a mechanism, AID expression in *Escherichia coli* caused a mutator phenotype and a C/G -> T/A transition mutation bias, which was exaggerated in the absence of the uracil excision enzyme, uracil DNA glycosylase (UDG). The generally accepted interpretation of these data was that AID could deaminate cytosines in DNA and the resulting non-repaired uracils (which base-pair like thymines) could template the incorporation of adenines and ultimately manifest as C/G -> T/A transition mutations.

The importance of uracil

The potent DNA cytosine deaminase activity of AID has now been demonstrated in a variety of model genetic and biochemical systems. The notion that it is this activity that directly triggers antibody gene diversification is supported by the fact that the mutational preference of purified AID (i.e. its intrinsic activity) is similar to that observed *in vivo* in the antibody gene variable or switch regions: AID prefers to deaminate DNA cytosines preceded by an A/T base pair and a purine (i.e. 5'-WRC).

A key prediction of the original DNA deamination model is that uracil is a central intermediate. The chicken B lymphocyte line DT40 can diversify immunoglobulin variable region DNA by IGC and also by SHM in some recombination-defective backgrounds. Normally DNA-based uracils are removed and ultimately repaired by UDG. However, when this enzyme was specifically inhibited in DT40, a dramatic bias toward C/G -> T/A SHM patterns resulted. A similar shift was observed in the

SHM patterns of mice lacking the major UDG (UNG2). Such a mutational shift is consistent with the interpretation that all types of base substitutions can result from error-prone DNA synthesis over an abasic site (created by uracil excision), whereas transitions at C/G base pairs result simply from DNA synthesis over non-excised uracils. Moreover, subsequent studies showed that levels of CSR in mice and humans and of IGC in DT40 were strongly reduced by the UDG deficiency. Together, these data indicate that uracil excision is a crucial step in all three of these processes.

A second prominent DNA repair system is also capable of recognizing U/G mispairs. Mismatch repair (MMR) normally functions to remove post-replicative lesions from DNA and, for more than a decade, has had a controversial association with SHM and CSR. However, recent experiments from the Neuberger group and others have clearly shown that the MSH2 protein is required, together with UDG, for aspects of SHM and for CSR. The CSR result is clear-cut, with both the UNG2- and the MSH2-deficient mice displaying lower levels of CSR and with the double mutant mice completely lacking non-IgM antibodies. The SHM result is more intriguing, with the UNG2-deficient mice showing a C/G -> T/A mutation bias and the MSH2-deficient (and the double mutant) mice showing dramatically fewer base substitution mutations at A/T bases (normally, base changes at A/T exceed those at C/G in humans and mice). This connection of mismatch repair to the A/T-directed SHM program has not been resolved mechanistically (although it is indeed also AID-dependent).

Directing AID

Now that considerable progress has been made towards understanding the mechanism of antibody gene diversification, a renewed attention is focusing on the question of how AID is targeted specifically to the immunoglobulin locus. A prominent feature of SHM and IGC is that the events occur specifically within the expressed variable region, not even upstream of the promoter or downstream in the constant region (let alone in other expressed genes in B cells). Given the fact that AID is predominantly a cytoplasmic protein, how does it gain access to the antibody gene DNA whilst shuttling in and out of the nucleus (Figure 1)? Moreover, when doing so, how does AID avoid mutating the bulk of the genomic DNA?

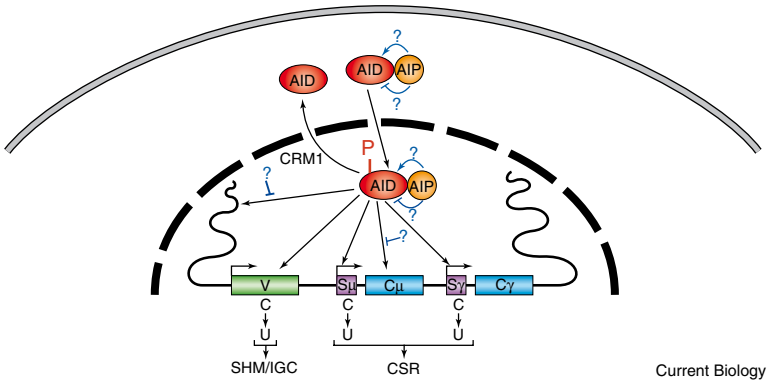

Figure 1. Antibody gene diversification triggered by AID.

AID has a predominantly cytoplasmic localization and must therefore diffuse or actively transport into the nucleus by an as-yet-undefined mechanism. Once inside the nucleus, AID appears to be directed toward the immunoglobulin variable and switch regions (e.g. Sμ), and concomitantly away from the immunoglobulin constant regions (e.g. Cμ) and the rest of the genome. This entire program is probably guided by AID-interacting proteins (AIPs). AIPs identified to date include DNA-PKcs, MDM2, PKA, RPA (32 kDa subunit) and RNA polymerase II. The PKA-dependent phosphorylation of AID is depicted. CRM1 escorts AID back to the cytoplasm. SHM, somatic hypermutation; IGC, immunoglobulin gene conversion; CSR, class switch recombination.

Work from the Honjo and Nussenzweig groups has revealed that the carboxyl terminus of AID is required for nuclear export and, surprisingly, is essential for CSR (but not for SHM/IGC). In contrast to full-length AID, which predominantly localizes to the cytoplasm of B cells, AID lacking its last 10 amino acids appears to distribute throughout the cell. Moreover, incubating cells with the CRM1/Exportin1 nuclear export inhibitor leptomycin B caused a similar phenotype. Therefore, the export of AID from the nucleus is most likely mediated by the CRM1 pathway.

Further studies with this interesting AID carboxy-terminal deletion mutant revealed that this region is dispensable for SHM/IGC but not for CSR. A closer analysis of the switch region DNA from B cells expressing this AID variant showed an abundance of AID-dependent base substitutions. One provocative interpretation of these data is that the carboxy-terminal domain of AID helps recruit a CSR-specific factor, which is required for the isotype switch mechanism, and that an SHM- (and/or IGC-) specific factor also exists. Indeed, some support for the latter possibility has come from studies of additional AID mutants suggesting that an amino-terminal region harbors SHM/IGC activity.

So, if CRM1 helps AID exit the nucleus, then one wonders how AID enters the nucleus and, more specifically, how it finds the expressed immunoglobulin genes. Recent studies from the Alt laboratory have shown that AID can interact with the 32 kDa subunit of the single-stranded DNA-binding protein complex, RPA. Biochemical data indicate that the binding of AID to RPA facilitates the deamination of transcribed WRC-rich immunoglobulin gene DNA. This is indeed plausible, but hard to rationalize as the sole AID-targeting mechanism, because many transcription and DNA replication, repair and recombination proteins have been shown to associate with RPA. Further work is necessary to explain, for instance, why AID is not attracted to RPA-bound DNA during replication, and to

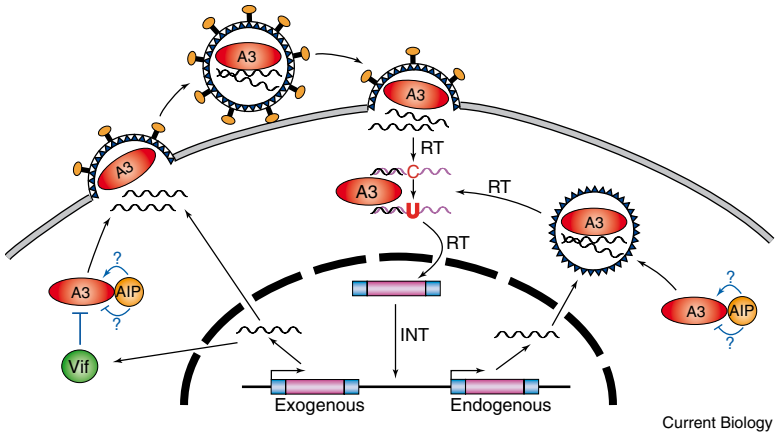

Figure 2. Retroelement restriction triggered by APOBEC3 proteins.

Endogenous and exogenous retroelement RNA is transcribed from elements integrated in the host genome and packaged into assembling particles. APOBEC3 (A3) proteins are incorporated into the particles (most likely through association with the viral Gag protein) and, upon reverse transcription (RT) of the viral RNA, they deaminate cytosines to uracils in the first strand cDNA. These events are followed by integration (INT) of the replicated retroelement DNA into the host cell's genome. HIV Vif targets APOBEC3F/G for proteasomal degradation. APOBEC3-interacting proteins (AIPs) include the viral proteins Gag and Vif, and the cellular ubiquitin ligation complex CUL5/ELOB/ELOC/RBX.

delineate the proteins that assist its nuclear entry.

The data from the Alt laboratory, in combination with a recent report from the Dalla-Favera group, also indicate that the nuclear-localized AID protein may be preferentially phosphorylated and that this specific phosphorylation is mediated by protein kinase A, which would neatly connect this event to cell signals that trigger antibody maturation. Nevertheless, other, as-yet-unidentified proteins are likely to be required for the proper targeting of AID to the antibody genes (Figure 1).

APOBEC3 DNA deaminases

AID is only one of at least ten putative or proven DNA cytosine deaminases encoded by the human genome. All of these proteins harbor a zinc-binding, cytosine deamination domain — $HxE–x_{23–28}–PCx_{2–4}C$ — in which the histidine and the cysteines coordinate zinc and the glutamate participates in deamination. AID and APOBEC1, the founding member of this family, are encoded by chromosome 12, APOBEC2 by chromosome 6 and APOBEC3A–3H by chromosome 22. Two members of the latter cluster, APOBEC3F and APOBEC3G, are cellular factors that can directly inhibit HIV infection.

Retrovirus restriction

APOBEC3G was identified as a dominant cellular factor that inhibited the replication of HIV variants lacking a small accessory protein called Vif. Several groups simultaneously showed that this 'restrictive' function of APOBEC3G was attributable to its entering newly synthesized virus particles, hitch-hiking with the particle to a new target cell and, during reverse transcription of the viral RNA genome, deaminating the nascent viral cDNA cytosines (Figure 2). The resulting uracils then template the incorporation of adenines, and a subsequent round of replication (or repair) accounts for the phenomenon that had been previously termed retroviral G -> A hypermutation. APOBEC3F has similarly potent anti-HIV activities and the combined mutation preferences of APOBEC3F and APOBEC3G account for the HIV hypermutation patterns observed in viral DNA sequences from AIDS patients.

If such a powerful defense mechanism operates *in vivo*, then how can HIV possibly replicate in humans and cause AIDS? Numerous groups independently showed that HIV Vif triggers the poly-ubiquitination and proteasome-dependent degradation of APOBEC3F and APOBEC3G. Biochemical studies from the Yu and Gabuzda groups have demonstrated the involvement of the CUL5-ELOB/C-RBX ubiquitin ligase complex. Vif thereby essentially disarms the cell in which HIV is being produced. It is notable that other cellular (and viral) proteins are marked for degradation by ubiquitin, leading to speculation that other AID/APOBEC family members might be similarly regulated and perhaps even share a subset of interacting proteins.

An evolutionary conundrum

Obviously, inhibition of HIV, though important, is not the evolutionarily preserved function of the human APOBEC3 proteins. A number of facts strongly suggest that their true reason for existence may be to control endogenous retroelements. First, it is widely accepted that HIV entered the human population within the past hundred years, whereas the APOBEC3 proteins are approximately 100 million years old. Second, a comparison of primate and human APOBEC3 protein sequences has indicated that they are almost all under a strong positive selective pressure(s) (i.e. the number of coding base changes far exceeds the number of silent base changes), suggesting the presence of a strong and probably ongoing genetic conflict. Third, recent work from several laboratories has shown quite convincingly that multiple APOBEC3 proteins, including APOBEC3F and APOBEC3G, are able to inhibit a broad spectrum of endogenous retroelements, such as the yeast retrotransposon Ty1 and the murine IAP and MusD retroelements. Finally, the infective potential of many distantly related retroviruses can be blocked by a single APOBEC3 protein. For instance, APOBEC3G can restrict murine leukemia virus, equine infectious anemia virus, simian immunodeficiency virus, feline and simian foamy virus, human T lymphotrophic virus, human hepatitis B virus and, as previously

described, HIV. Thus, all of these observations combine to indicate that the restrictive potential of the APOBEC3 proteins is broadly applicable.

One possible mechanistic link is that many of the APOBEC3 proteins seem to exploit the multifunctional and ubiquitous Gag protein in order to access the single-stranded DNA of replicating retroelements. Indeed, an association has been reported between APOBEC3G and the distantly related Ty1 and HIV Gag proteins. However, it is important to point out that Gag associates with RNA and that APOBEC3 proteins can associate with RNA, so their mutual association may also be RNA mediated (a controversial issue in the literature). Structure–function studies are desperately needed to help shed additional light on this important area.

Directing APOBEC3

Like AID, APOBEC3F and APOBEC3G appear to have a predominantly cytoplasmic subcellular localization (Figure 2). This location seems like a sensible way to gain access to the cores of assembling retroelements, as well as keep the potentially hazardous mutator activity safely away from the genomic DNA of the cell. However, expression of APOBEC3G in yeast was recently shown to trigger genomic hypermutation. This implies that the DNA cytosine deaminase activity of APOBEC3G can penetrate both the eukaryotic nuclear membrane (omnipresent in *S. cerevisiae*) and chromatin. Thus, it is reasonable to postulate that, like AID, the human APOBEC3 proteins are tightly regulated. Recent studies from Greene's laboratory support this contention, as they indicate that APOBEC3G can associate with a near mega-dalton cellular complex. It is likely that APOBEC3F, which appears to be co-expressed and can interact with APOBEC3G, will also associate with the same complex. Additional work is necessary to define these and likely other cellular regulators of the human APOBEC3 proteins.

Common origins

AID has a pivotal role in generating antibody diversity, and APOBEC3F/G clearly possesses the capacity to block retroelement replication. DNA cytosine deamination is key to both mechanisms. Although many of the other human APOBEC3 proteins are still less well understood, it is likely that they may also have roles fortifying the innate immune defenses against pathogens lurking within our systems. Phylogenetic studies have suggested that the APOBEC3 proteins evolved from an AID-like ancestor. Taking this together with the fact that non-mammalian vertebrates lack APOBEC3 proteins, one might wonder whether the AID protein from such organisms might very well possess both antibody diversification and retroelement restriction activities? Thus, ancestral AID may have possessed retroelement restriction activities that were subsequently co-opted for antibody diversification and other biological processes. In other words, is AID the mother of all APOBECs?

Acknowledgements

The important contributions of many investigators could not be specifically cited due to space considerations, but many of these can be found within the bibliography. We thank Tim Behrens, Bill Brown, Eric Hendrickson, Stéfan Jónsson and Michael Neuberger for thoughtful comments on this manuscript. R.S.H. is supported in part by a Searle Scholarship, a University of Minnesota McKnight Land Grant Assistant Professorship, and a grant from the NIH NIAID (AI064046).

Further reading

Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J.P., Manis, J. P., and Alt, F.W. (2005). The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature *438*, 508–511.

Bransteitter, R., Pham, P., Calabrese, P., and Goodman, M.F. (2004). Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J. Biol. Chem. *279*, 51612–51621.

Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature *430*, 992–998.

Chiu, Y.L., Soros, V.B., Kreisberg, J.F., Stopak, K., Yonemoto, W., and Greene, W.C. (2005). Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. Nature *435*, 108–114.

Delebecque, F., Suspene, R., Calattini, S., Casartelli, N., Saib, A., Froment, A., Wain-Hobson, S., Gessain, A., Vartanian, J. P., and Schwartz, O. (2006). Restriction of foamy viruses by APOBEC cytidine deaminases. J. Virol. *80*, 605–614.

Dutko, J.A., Schafer, A., Kenny, A.E., Cullen, B.R., and Curcio, M.J. (2005). Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. Curr. Biol. *15*, 661–666.

Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A. J., Heidmann, T., and Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature *433*, 430–433.

McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P., and Nussenzweig, M.C. (2004). Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. J. Exp. Med. *199*, 1235–1244.

Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006). PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. Proc. Natl. Acad .Sci. USA *103*, 395–400.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99–103.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol. Cell *16*, 163–171.

Sawyer, S.L., Emerman, M., and Malik, H.S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol. *2*, E275.

Schumacher, A.J., Nissley, D.V., and Harris, R.S. (2005). APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. Proc. Natl. Acad. Sci. USA *102*, 9854–9859.

Shinkura, R., Ito, S., Begum, N.A., Nagaoka, H., Muramatsu, M., Kinoshita, K., Sakakibara, Y., Hijikata, H., and Honjo, T. (2004). Separate domains of AID are required for somatic hypermutation and class-switch recombination. Nat. Immunol. *5*, 707–712.

Shirakawa, K., Takaori-Kondo, A., Kobayashi, M., Tomonaga, M., Izumi, T., Fukunaga, K., Sasada, A., Abudu, A., Miyauchi, Y., Akari, H., *et al.* (2006). Ubiquitination of APOBEC3 proteins by the Vif-Cullin5-ElonginB-ElonginC complex. Virology *344*, 263–266.

Simon, V., Zennou, V., Murray, D., Huang, Y., Ho, D. D., and Bieniasz, P. D. (2005). Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. PLoS Pathog. *1*, e6.

Suspene, R., Guetard, D., Henry, M., Sommer, P., Wain-Hobson, S., and Vartanian, J. P. (2005). Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. Proc. Natl. Acad. Sci. USA *102*, 8321–8326.

University of Minnesota, Department of Biochemistry, Molecular Biology and Biophysics, Institute for Molecular Virology, Beckman Center for Transposon Research, Minneapolis, Minnesota 55455, USA.
E-mail: rsh@umn.edu